DOCUMENT RESUME

ED 222 542 TM 820 694

AUTHOR Gustafsson, Jan-Eric
TITLE New Models of the Structure of Intellectual Abilities: Implications for Testing and Teaching Practice.
INSTITUTION Goteborg Univ., Molndal (Sweden). Dept. of Education.
SPONS AGENCY National Swedish Board of Education, Stockholm.; Swedish Council for Research in the Humanities and Social Sciences, Stockholm.
PUB DATE Jul 82
NOTE 49p.; Paper presented at the International Congress of Applied Psychology (20th, Edinburgh, Scotland; July 25-31, 1982).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cognitive Ability; Educational Testing; Factor Analysis; *Factor Structure; Intermediate Grades; *Models; *Testing
IDENTIFIERS *Confirmatory Factor Analysis; *Hierarchical Analysis

ABSTRACT

The problem of the structure of human abilities is approached within the framework of higher order confirmatory factor analysis. Non-hierarchical models suggested by Thurstone and Guilford are reviewed, and it is concluded that these models fail to give a theoretically and practically useful representation of the organization of human abilities. Hierarchical models proposed by Vernon and Cattell-Horn are compared, and it is concluded that the difference between these models may be more apparent than real. It is also concluded that it may be possible to construct a unified hierarchical model, which includes ability dimensions of at least three different levels of generality. This hypothesized model is tested in two empirical studies, which both yield support for the model. Implications of the model for testing are discussed and it is concluded that while the model is compatible with much established testing practice, it goes beyond this by allowing hierarchically differentiated assessments. Included is a demonstration of how the model may be used in research on aptitude-treatment interactions, by simultaneous analysis of factors at different levels of generality in structural equation models. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED222542

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Gustafsson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

1982:03

New models of the structure of intellectual abilities: Implications for testing and teaching practice

Jan-Eric Gustafsson

Paper presented at the 20th International Congress of Applied Psychology, Edinburgh, 25th - 31st July 1982.
The research presented in the paper has been supported financially by the National Board of Education in Sweden, and by the Swedish Council for Research in the Humanities and Social Sciences.

Table of Contents

1 INTRODUCTION.....1

2 PREVIOUS MODELS.....2
2.1 The Spearman theory of Two Factors.....3
2.2 Multiple Factor models.....4
2.3 Hierarchical models.....5
2.4 Discussion.....7

3 HIERARCHICAL, LISREL-BASED, MODELS.....10
3.1 Study I: Reanalysis of Undheim (1978).....10
3.2 Study II: Analysis of a test-battery for Gf, Gc and Gv.....11
3.3 Discussion.....14
3.3.1 Interpretations of the factors.....14
3.3.2 Relationships with other models.....18
3.3.3 The nature of the HILI-model.....22

4 IMPLICATIONS OF THE MODEL.....22
4.1 Implications for testing.....23
4.2 Implications for teaching.....25
4.2.1 An example of the HILI-model in ATI-research.....27
4.2.2 Results.....29
4.2.3 Discussion of the results.....30

5 CONCLUDING REMARKS.....31

ABSTRACT

In the paper the problem of the structure of human abilities is approached within the framework of higher-order confirmatory factor analysis. Non-hierarchical models suggested by Thurstone and Guilford are reviewed, and it is concluded that these models fail to give a theoretically and practically useful representation of the organization of human abilities. Hierarchical models proposed by Vernon and Cattell-Horn are compared and it is concluded that the difference between these models may be more apparent than real. It is also concluded that it may be possible to construct a unified hierarchical model, which includes ability dimensions of at least three different levels of generality. This hypothesized model is tested in two empirical studies, which both yield good support for the model. Implications of the model for testing are discussed and it is concluded that while the model is compatible with much established testing practice, it goes beyond this by allowing hierarchically differentiated assessments. It is also demonstrated how the model may be used in research on aptitude-treatment interactions, by simultaneous analysis of factors at different levels of generality in structural equation models.

1 INTRODUCTION

Research on the structure of abilities aims to study which dimensions of intelligence may be identified, and how these are related to one another. Such research has been conducted, most often by means of factor analysis, since the beginning of the twentieth century, with major contributions by, among others, Spearman (1904b), Thurstone (1938), Vernon (1950), Guilford (1967), Horn (1968), and Cattell (1971).

Development within the field has not been smooth and gradual, however. Rather, it has been characterized by rather intense struggles between different schools of thought, and periods of intensive research have been interspersed with periods of lighter activity. Nor does it seem that there now exists a commonly accepted answer to the question how human abilities are organized.

Sternberg (1981b) argued that the psychometric line of research on intelligence has failed because exploratory factor analysis has been so successful in supporting many alternative models of intelligence. Whatever is the reason, however, there is little doubt that the lack of a more commonly accepted model of the structure of abilities has hampered development of both theory, applications, and other lines of research, such as research on aptitude-treatment interactions (Cronbach & Snow, 1977).

During the last decade or so the focus in research on intelligence has changed from questions about structure, to questions about the processes and components of intelligence (e.g. Curtis & Glaser, 1981; Sternberg, 1977). It is, of course, much too early to evaluate the results of these newer approaches to the understanding of intelligence. It would seem, however, that the lack of a commonly accepted model of the structure of abilities may hamper this research as well. Thus, one obvious consequence of the lack of a common frame of reference for the classification of abilities is that different researchers employ quite different labels and interpretations for quite similar sources of individual differences. This will make it difficult, if not impossible, to integrate the results achieved.

There seem, thus, to be several reasons to make a renewed attack on the problem of the structure of human abilities. The present paper makes such an attempt, using techniques of confirmatory, rather than exploratory, factor analysis to study the organization of cognitive abilities. The layout of the paper is such that some of the dominating models of the structure of abilities are first reviwed, then two empirical studies are presented along with a proposed new model (the HILI-model), and finally some implications of the model are discussed with a focus on testing and ATI-research.

2 PREVIOUS MODELS

In late nineteenth century several attempts were made to measure intelligence through psychophysical assessments (e.g. Galton, 1883). These attempts were not successful, however, (e.g. Wissler, 1901), and it was not until Binet and Simon at the turn of the century chose to concentrate upon more complex tasks that individual differences in intelligence could be satisfactorily scaled.

The Binet and Simon test consisted of a wide assortment of tasks, chosen on the basis of their ability to discriminate between different age groups. No theory indicated, however, why the items had such capacity. But because of the practical utility of the test it got followers and it established a tradition of measuring intelligene with heterogenous collections of tasks.

Ever since Binet and Simon found a way of measuring intelligence attempts have been made to define the concept. Unanimosity has not been reached (e.g. Resnick, 1976), however, and it is still under debate whether it is at all meaningful to identify a concept such as general intelligence. Binet himself, for example, did not conceive of intelligence as a separate capacity but viewed it as comprised of several more or less clearly identifiable capacities, such as judgement, common sense, initiative and ability to adapt. The ability measured by the test would thus be some kind of average of the different capacities involved.

This problem of the dimensionality of intelligence has been in focus in the factor-analytically based differential psychological research, which took off at about the time when Binet and Simon published their test. As has already been mentioned it has not been possible to derive one single, generally agreed upon, model of the structure of abilities. Instead different schools have been formed, each associated with a particular technique of exploratory factor analysis, from which has followed more or less directly a certain model. Some of these models are now of historic interest only, but several have survived into our days. It may, therefore, be worthwhile to describe in some detail the most influential models.

2.1 The Spearman theory of Two Factors

Binet attempted to measure "general intelligence" even though his theory of intelligence did not include such a concept. Spearman (1904b, 1927), among others, noted this lack of compatibility between theory and observational technique and argued that if the general ability measured by the test is an average or sample of different abilities, its nature is determined by what items happen to be included in the test: "No genuine averaging, or sampling of anybody's abilities is made, can be made, or even has really been attempted " (Spearman, 1927, p. 71, emphasis in original).

Spearman (1904a) developed, however, a technique to show whether one common factor is sufficient to account for the intercorrelations among a set of tasks. Applying this technique to several different sets of variables, comprising among other things psychophysical assessments, ratings of intelligence, school marks, and psychological tests, he was able to conclude (Spearman, 1904b, 1927) that in most cases one underlying factor accounted for the intercorrelations. On the basis of these results he proposed the Two Factor theory of intelligence, which states that performance on an intellectual task is affected by two factors only: one general (g) and one specific (s). The g factor enters more or less prominently into any intellectual activity, and the individual's standing on this factor is the same irrespective of the task. The s factor is of little importance when the g factor is of great importance and vice versa. However, for each type of intellectual activity a different s factor is assumed, so the individual's standing on this factor varies with the task.

On the basis of this model Spearman concluded, among other things, that the g factor may indeed be measured with a heterogeneous collection of tasks. This is due to the fact that each task contributes information on the g factor, while the s-factors cancel ("the principle of indifference of the indicator"). In empirical studies he found, however, the involvement of g to be largest in tasks demanding "eduction of relations and correlates", as in tests of analogical reasoning, or in tests with series items.

As has already been mentioned Spearman was able to secure empirical support for his Two Factor theory. The fit of data to the model was far from perfect, however. In particular it was found that when variables which were "too similar" were analyzed together, the model broke down because of correlations between s factors. Spearman (1927) recommended that the model itself should decide when two tests are "too similar", but this left the theory open to the same kind of criticism of arbitrariness as Spearman had leveled against Binet, and the Two Factor theory was soon thoroughly refuted (e.g. Kelley, 1928).

2.2 Multiple Factor models

Spearman (1904a) invented the first factor analytic model. The limitations inherent in the fact that the technique only handles one factor became obvious, however, and more refined techniques capable of representing systematic variance of much greater complexity were developed.

The major contributor was Thurstone (1931; 1938; 1947), who developed factor analysis to encompass multiple common factors. Applying the new technique to large test batteries (e.g Thurstone, 1938), about a dozen factors could be identified. By performing rotations according to the principle of simple structure, which essentially states that any test should be affected by one or a few factors only, it was found that each factor accounted for performance on a subset only of the tests in the battery. There was no sign of a general factor.

Most factors identified by Thurstone (1938) were replicated by Thurstone and his colleagues (e.g Thurstone, 1940; Thurstone & Thurstone, 1941), and it was possible to set up a list of six or seven easily replicable Primary Mental Abilities (PMA´s), such as Verbal Comprehension (V), Word Fluency (W), Induction (I), Space (S), and Perceptual Speed (P).

In the first set of analyses the PMA´s were kept orthogonal. But when test batteries were assembled to measure the factors, it was found that the tests were intercorrelated, which led Thurstone to adopt instead an oblique factor model.

The techniques of factor extraction and rotation invented by Thurstone have by now been considerably refined (Harman, 1967). The basic principles remain the same, however, and the Thurstonian Multiple Factor analysis has evolved into the dominating factor analytic technqiue.

The basic PMA model, with a handful of factors, has been elaborated upon too. By investigation of new domains of ability, and by showing that several of the original PMA´s are differentiable into more narrow factors, the list of factors has been extended considerably. Reviews of the research (e.g. French, 1951; French, Ekstrom & Price, 1963; Horn, 1977; Pawlik, 1966) indicate that it may be necessary to distinguish between as many as 30 to 50 factors of ability. While some of these factors are broad and comprehensive, others are very narrow and circumscribed, and must be considered the result of a subdivision of a broader factor. However, all factors are primary in the sense that they represent results from applications of Multiple Factor analysis to matrices of intercorrelations between tests.

The proliferation of factors has been carried even further in the Guilford "Structure-of-Intellect" (SI) model (e.g. Guilford, 1967). This model describes tests and factors in terms of the three facets operation, content and product. Combination of the facets yields no less than 120 possible factors, and in the latest version of the model Guilford and Hoepfner (1971) claim identification of at least one factor (and sometimes more) in each of 98 of the cells in the model.

In spite of the fact that the SI-model is a morphological model structured along three dimensions, it is assumed that the factors are uncorrelated. This in turn implies an assumption that factors having one or two facets in common are no more related than are factors which have no facets in common.

Multiple Factor analysis was introduced because of an obvious need to differentiate between more factors than was allowed by the Spearman technique. However, by now so many factors have been identified that the results appear almost incomprehensible and of limited practical utility (e.g. Humphreys, 1962; McNemar, 1964). Undheim (1981c) even went so far as to argue that:

> The widespread application of multiple factor analysis in research on abilities seems to have carried factor analysis far beyond its descriptive and conceptual limitations as a research tool, resulting in an ever increasing number of factors of "the mind". It is somewhat paradoxical that whereas the multiple-factor model developed by Guilford ... tried explicitly to save ability research from this empirical chaos, it, in fact, may have contributed to further "empiricism" in focusing on filling the empty cells in the box model of 120 factors. (Undheim, 1981c, p. 22)

We will return to the problem of proliferation of factors later on. Before that, however, there is reason to describe another approach to understanding the structure of abilities.

2.3 Hierarchical models

In the British research on abilities Multiple Factor analysis has had less impact, and the g factor has not vanished altogether. In the post-Spearman work it was soon to be discovered, however, that in addition to g there are also group-factors of great importance. Factor analytic techniques were developed, which from a matrix of correlations extract first the g factor, and then group-factors of successively smaller breadth.

These hierarchical group-factor techniques (e.g Burt 1941; Harman, 1967) thus have the advantage of being able to supply information both about a general factor and about group-factors.

One hierarchical model, summarizing the results of very many analyses, was proposed by Burt (1949). This model seems, however, to have been too much of a logically constructed classification scheme to earn any great impact. Instead a rather similar model, presented only slightly later by Vernon (1950, 1965) has received more widespread attention.

At the top of the Vernon model (see Figure 1) there is the g factor, and at the next level below there are two major group-factors: verbal-educational (v:ed) and spatial-practical-mechanical (k:m) ability. The v:ed factor subdivides into different scholastic factors, such as number factors and linguistic abilities. The k:m factor subdivides too and this complex includes minor group-factors such as perceptual, spatial and mechanical factors. Each of these minor factors can then be subdivided by more detailed testing.

Insert Figure 1 about here

Another hierarchical model has been proposed by Cattell and Horn. The basic concepts of this model were developed by Cattell (e.g. 1940), but the model was neither elaborated upon, nor put to empirical tests until considerably later (e.g. Cattell, 1963; Horn, 1965, 1968; Horn & Cattell, 1966).

Methodologically the Cattell-Horn model is based upon oblique Multiple Factor analysis of several orders. Thus, in the first step an ordinary oblique Multiple Factor analysis is conducted, typically yielding a large set of primary or first-order factors. The correlations between the primary factors are then subjected to another factor-analysis which yields secondary or second-order factors. In principle this procedure of factoring may be continued at successively higher orders. Cattell and Horn have, however, chosen to stop the factoring at the second-order level.

In the Horn and Cattell (1966) formulation the model includes 5 second-order or "general" factors. The two most important ones are fluid intelligence (Gf) and crystallized intelligence (Gc). Both are viewed as aspects of general intelligence and are said to involve abstraction, concept formation, and perception and eduction of relations. Gf, however, is involved in tasks which are new to the examinee, while Gc is shown in tasks with verbal-conceptual content. Gf is thought to represent influences of biological factors and

incidental learning on intellectual development, while Gc is interpreted as reflecting education and experience. Primary factors such Induction (I) and Cognition of Figural Relations (CFR) involve Gf, while Gc has been found to be involved in primaries such as Verbal Comprehension (V) and Cognition of Semantic Relations (CMR).

The other three second-order factors in the Cattell-Horn model are General visualization (Gv), General fluency (F or Gr), and General speediness (Gs). Gv is involved in almost all tasks with figural content and runs strongly through primaries such as Visualization (Vz), Flexibility of Closure (Cf), Speed of Closure (Cs) and Spatial Orientation (S). Gr reflects the flexibility with which labels for cultural concepts are recalled and recognized, and is involved in, among others, the primary factors Associational fluency (Fa), Word Fluency (Fw), and Ideational Fluency (Fi). Gs is defined as quickness of performance, and shows up most clearly in primary factors defined by very simple tasks, such as Perceptual Speed (P) and Numerical Facility (N).

The model has been subjected to empirical tests in large scale factor analytic studies by among others Cattell (1963), Horn (1972), Horn and Cattell (1966) and Undheim (1976, 1978, 1981a). These studies have in general confirmed the hypothesized structure.

It should also be pointed out that the basic model as described above has been elaborated upon by Cattell (1971) and Horn (in press). It would seem, however, that the model in the 1966 formulation so far has had greatest impact and has the best empirical support.

2.4 Discussion

Among the models discussed above, those based upon Multiple Factor analysis have dominated differential psychology during the last three or four decades. One reason for this is that they do show a good fit to empirical data; another reason may be the ready availability of computer programs to perform Multiple Factor analysis.

As has already been pointed out these models give, however, a rather fragmented picture of the structure of human abilities (e.g. Humphreys, 1962; Undheim, 1981c), and the practical utility of distinguishing a very large number of ability factors has been questioned (e.g. McNemar, 1964; Cronbach & Snow, 1977, pp. 152-164).

1981c), in particular, has gone far towards a "restoration of general intelligence" on the basis of a reinterpretation of Gf in terms of g. Even though Undheim has not yet presented any conclusive empirical proof that this is the case, his line of reasoning provides additional support for the hypothesis.

The factor analytic techniques employed in previous research tend to be biassed in favor of one of the models. These techniques also are fraught with the problem that they are exploratory only, and do not give provisions for statistically sound tests of the number of common factors, or of the significance of factor loadings. Recently, however, factor analytic methods have been developed which do allow the testing of hypotheses, and which are flexible enough to allow an almost infinite range of different models to be specified.

Jöreskog (1969) presented a method for estimating and testing confirmatory factor models, using maximum likelihood methods. In such models the number of factors, and the pattern of loadings is specified in advance, on the basis of whatever previous knowledge is available about the variables being measured. Estimates of parameters in confirmatory models are unique, so the problem of rotation is avoided altogether. Statistical tests also are available with which the fit of the data to the model may be determined.

Jöreskog (1970) generalized the simple confirmatory factor analytic model to allow formulation of higher-order models, and in still further developments a model has been arrived at which, loosely stated, combines the factor analytic methods with path-analytic techniques (linear structural relations, LISREL; Jöreskog, 1973; Jöreskog & Sörbom, 1978, 1981). This latter model is a completely general model which contains all the earlier models as special cases.

It would thus seem that major progress has been made in estimating and testing hierarchical models. This technolopgy has been put to use in a series of studies, the results of which are summarized below.

The problems associated with the multitude of factors are severed by the fact that Multiple Factor analysis, in a technical sense at least, considers all factors as being of equal importance. In the hierarchical models, in contrast, the lower-order factors are subsumed under higher-order factors, which makes these models parsimonious, while they at the same time provide the richness of description of the Multiple Factor models. Thus, in the Cattell-Horn model the Thurstonian primaries, along with several of the factors in the SI-model, are found at the first-order level. This model, therefore, is compatible with the Multiple Factor models, but it obviously goes much beyond these models.

It does appear that during the last few years the hierarchical type of model has gained increasing attention (e.g. Cronbach & Snow, 1977; Snow, 1977, 1980). Given the advantages of this type of model it may be asked, however, why the influence has not been even greater. The most important reason for this is most likely that they are technically quite complicated to deal with. Another reason, however, may be that there are several competing models, which seem so different that if one of them is correct, the others must necessarily be refuted.

The most important hierarchical models are the Cattell-Horn and Vernon models, and Cattell (1963) and Horn (1968), in particular, have claimed the superiority of their model. They argue that Vernon's v:ed should correspond to Gc, and that k:m corresponds to a mixture of Gf and Gv. In their view, then, a major difference between the models is that where Vernon distinguishes two factors, they distinguish three. A similar point was made by Humphreys (1967), who argued that the answer to the question whether Gf and Gv are distinct factors determines whether the Cattell-Horn or the Vernon model should be accepted.

However, another set of relationships between the models may also be hypothesized. The kind of tests identified to measure Gf comes very close to the kind of tests that Vernon lists as measures of g in his model. It may, therefore, be argued that Gf in the Cattell-Horn model is more or less the same factor as the g factor in the Vernon model (cf. Vernon, 1969, p. 25). If this is true, it also would seem natural to equate Gv with k:m, and Gc with v:ed. This implies that if a third level, representing the g factor, is added to the Cattell-Horn model, the most essential point of difference between the two major hierarchical models would be resolved. Such a combined hierarchical model would also be compatible with most of the non-hierarchical models described above. It is, of course, an empirical question whether the second-order factor labeled Gf by Cattell and Horn is indeed the same factor as the one called g by Spearman and Vernon, and it is also an empirical question whether the other parts of this hypothesized model fit empirical data. It may be noted, however, that Undheim (1981b,

3 HIERARCHICAL, LISREL-BASED, MODELS

It has been concluded that the difference between the Cattell-Horn and Vernon models may be smaller than is evident at first sight, and it has also been suggested that it may be possible to construct another model, as a synthesis of these two models. For practical reasons it is quite impossible to test this model in one single empirical study. Several studies have, therefore, been performed, each of which tests partly different aspects of the model. Below the results of two such studies are briefly presented (for a fuller account, see Gustafsson, Lindström & Björck-Åkesson, 1981).

3.1 Study I: Reanalysis of Undheim (1978)

Undheim (1978) administered a test battery of 30 tests to a sample of 149 6th grade children, with the purpose of testing the Cattell-Horn model. Above all interest was centered on the question whether Gc and Gf are differentiable in a sample of 12-13 year old children.

The tests in the Undheim battery were hypothesized to measure some 12 different primary factors. However, some of the factors were represented by one test only, so an exploratory factor analysis could not aspire to identify all the primaries. Using principal factor analysis 5 factors were extracted, and rotated to simple structure with a variety of oblique and orthogonal methods of rotation. The rotated factors could be interpreted to represent Gf, Gc, Gv, Gs and Gr. Even though the factors were identified at the first-order level these results may be taken to provide good support for the Cattell-Horn model.

Through the courtesy of Dr. Undheim, the matrix of intercorrelations among the tests has been made available, and a reanalysis has been performed in which a series of higher-order LISREL models have been fitted.

The final model is shown in Figure 2. As may be seen in the Figure the model includes 9 first-order factors, almost all of which correspond to primary factors originally hypothesized by Undheim. It proved impossible, however, to make distinctions between the Induction (I) factor and a Figural Relations factor hypothesized by Undheim. Nor was it possible to differentiate between three different hypothesized primary factors in the Gv-domain, so Gv is identified as an undifferentiated primary factor.
There are three second-order factors in the model: Gf, defined by the Reasoning (R) and I factors; Gs, defined by Perceptual Speed (P) and Number

---

Insert Figure 2 about here

Facility (N); and Gr, defined primarily by Word Fluency (Fw) and Ideational Fluency (Fi).

Since there is one primary factor only in the Gc domain, it was impossible to define this factor at the second-order level. The first-order V-factor was, therefore, taken to represent the Gc-factor, and this factor, along with Gf, Gr, Gs and the first-order Gv, defines the third-order g factor. Most interestingly, the standardized loading of Gf in g was found to be 1.0, which result does support the hypothesis of an identity between Gf and G.

The model presented in Figure 2 has a very good fit (chi-square=370.9, df=329, $p < .06$), which indicates that it cannot be rejected as a proper representation of the data.

It may be noted that at the first- and second-order levels these results bring very strong support to the Cattell-Horn formulation, except that Gv was found to be an undifferentiated primary factor and not a second-order factor. A very large number of studies have, however, shown the Gv domain to contain several clearly identifiable primary factors (e.g. Lohman, 1979). It is likely, therefore, that the results of the present study are due to the fact that the Undheim sample consisted of a majority of young females, for which the differentiation of Gv has been found to be lesser (e.g. Smith, 1964).

3.2 Study II: Analysis of a test-battery for Gf, Gc and Gv

The second study to be summarized here was conducted by Gustafsson et al. (1981). Subjects in that study were some 1200 pupils in the 6th grade who were given a test battery consisting of 13 tests of cognitive ability and 3 standardized achievement tests. Of the total sample, 981 subjects had complete data on all the variables.

The tests in the battery were hypothesized to represent the following primary factors:

- Induction (I), measured by the tests Number Series (NS) and Letter

Grouping (LG). These tests were newly developed for the study, but were modelled upon tests previously constructed by Svensson (1971) and Härnqvist (1968).

- Cognition of Figural Relations (CFR), measured by the Raven Progressive Matrices Test. For the analysis the test was split into two half-tests, by assigning odd numbered items to one test, and even numbered items to the other (RA-O and RA-E).

- Memory span (Ms), measured by the Auditory Letter Span and Number Span tests from the ETS battery (French et al., 1963).

- Visualization (Vz), measured by the test Metal Folding (MF), constructed by Svensson (1971). For the analysis the MF test was split into half-tests, according to the odd/even rule to give the sub-tests MF-O and MF-E.

- Spatial Orientation (S), measured by the two parts of the Card Rotations (CR) test in the ETS battery (French, et al., 1963).

- Flexibility of Closure (Cf), measured by the Group Embedded Figures Test (GEFT, Witkin, Oltman, Raskin & Karp, 1971), and the Hidden Patterns (HP) and Copying (CO) tests from the ETS kit.

- Speed of Closure (Cs), measured by the tests Disguised Pictures (DP) and Disguised Words (DW). These tests were newly developed for the present study, but they were modelled upon already existing tests, such as Street and Mutilated Words (e.g. Thurstone, 1944).

- Verbal Comprehension (V), measured by the vocabulary test Opposites (Svensson, 1971). In the analysis of data the items in the test were split into two half-tests according to the odd/even rule (OP-O and OP-E).

- Scholastic Achievement (Ach), measured by Standardized Achievement test in Swedish (SA), English (EA), and Mathematics (MA).

A detailed description of each of the tests would carry too far in this context (see instead Gustafsson et al., 1981). They do all, however, come close to the tests of primary abilities developed by Thurstone and followers.

The hypothesized primary factors were in turn hypothesized to represent the three second-order factors Gf, Gc and Gv according to the following pattern: Gf was thought to be loaded by I, CFR, and to some extent by Ms; Gv was hypothesized to be loaded by Vz, S, Cf and Cs; and Gc was thought to be loaded by V and Ach.

These hypotheses were tested in a sequence of LISREL-models, in which successively the hypothesized first- and second- order patterns were fitted. The final model is shown in Figure 3.

Insert Figure 3 about here

This model fitted the data quite well (chi-square=184.5, df=144, p <.013, N=981), but to achieve this level of fit some modifications of the originally hypothesized model had to be introduced. Most of these modifications were minor ones, involving specification of covariances between the specific parts of tests and factors. Some of the modifications deserve mentioning, however. Thus, Cs was found to be more related to the Gf-factor than to Gv. This result should, however, not be taken as proof that this primary factor has been misclassified in previous research since the Cs factor in the present study, as is often the case, was only weakly identified. Furthermore, in the Gc-domain it was found that the hypothesized Ach-factor split into two (Verbal Achievement and Numeric Achievement), in spite of the dearth of tests to identify achievement factors in the present study.

So far the results are very much in accord with the Cattell-Horn model. In the next step of the analysis, however, a third level was introduced. Since there only are three second-order factors, a model with a third-order g factor is "just-identified", i.e. it has the same degrees of freedom and chi-square value as the second-order model. Restrictions may be imposed, however, to make the third-order model a testable one. Estimating the model under the constraint that g is identical with Gf did not worsen the fit significantly (chi-square=3.2, df=1). However, imposing the corresponding constraints with respect to Gv and Gc very significantly worsened the fit of the model (chi-square=98.2, df=1 and chi-square=126.5, df=1, respectively). These results thus strongly support the hypothesis that the second-order factors in the Cattell-Horn model are not really of the same order, but that Gf should be moved one step up in the hierarchical model.

3.3 Discussion

In Figure 4 the main results from the two empirical studies have been put together into a hierarchical model. This model will be referred to as the HILI-model, where the acronym stands for hierarchical, LISREL-based model.

Insert Figure 4 about here

The model has three levels, with a g factor at the top, the second-order factors of the Cattell-Horn model at the second level, and Thurstonian and Guilfordian primaries at the lowest level.

3.3.1 Interpretations of the factors

Both empirical studies showed clearly the g factor to be identical with Gf, which factor in turn has been found (e.g. Gustafsson et al. 1981) to be very highly loaded by the I-factor. Both Spearman (1927) and Vernon (1950) list the I-type of tests as the best measures of g, so there is little doubt that the g factor of the HILI-model is identical with the g factors of the Spearman and Vernon models. It is also interesting to note that when Thurstone and Thurstone (1941) performed a second-order analysis of the PMA's, they found one second-order factor (g), which turned out to be most highly loaded by the I-factor.

The g factor has been remarkably absent from theoretical discussions since the pioneering work of Spearman. At best it might be said that attention has now and then been called to the fact that the factor exists (e.g. McNemar, 1964; Humphreys, 1962). During the last few years, however, it appears that interest in the general factor has been revitalized, as may be seen in the writings of for example Humphreys (1979), Snow (1977, 1978, 1980), Sternberg (1980, 1981a) and Undheim (1981c).

As has already been mentioned, Undheim (1981a, 1981c) has proposed ideas almost identical with those espoused here, and from a series of studies it was concluded:

> Although Cattell's hypothesis of two intelligence factors, fluid

> and crystallized intelligence, is seemingly supported by the simple-structure factor analytic distinctions of two such factors in several studies, hierarchical order analysis indicate that these findings may support an alternative hierarchical model of intelligence where fluid tasks are central in the definition of general intelligence and group factors of crystallized ability or verbal-educational knowledge, visualization, and speediness emerge. Thus the results are consistent with a more parsimonious neo-Spearman structuring of broad ability factors (Undheim, 1981a, pp 184-185).

This conclusion thus conforms with the conclusion drawn here.

In his interpretation of general intelligence Undheim (1981c) stressed that g is a consequence of learning, and that the nature of intelligence is determined by cultural values: "general intelligence is good reasoning with the contents of our culture" (Undheim, 1981c, p. 256). This line of reasoning led Undheim to suggest a very broad definition of general intelligence, namely that it represents the entire repertoire of knowledge, skills and strategies. Undheim also concluded that:

> a measure of general intelligence should sample achievements in many subject matters -- some of which are tied to the academic curricula that subjects are exposed to, others tied to intellectual achievements acquired out of school (Undheim, 1981c, p. 257).

One problem with this "sampling" interpretation of g is that it disregards the fact that Gf has been shown to be identical with g. Formulated in simple terms this result implies that scores obtained on a test consisting of the broadest possible and most representative sample of tasks are virtually perfectly correlated with scores obtained on a small set of Gf tasks. The most interesting question must then be why the Gf-tests have such power of indexing general intelligence.

In speculations on the nature of g many authors have stressed that one important characteristic of tasks to measure g or Gf is that they present the examinee with new problems (e.g. Spearman, 1927, pp. 161-198; Horn, 1968; Cattell, 1971). Sternberg (1981a) has presented some empirical evidence that intelligence can best be understood through what he calls "nonentrenched" (i.e. novel) kinds of tasks.

Snow (1980) has gone one step further and has outlined a process theory of intelligence, in which theory too the novelty of tasks is seen as essential. Snow suggests that tests of general ability in particular may pose demands for new assembly of performance processes:

> Perhaps they represent to a greater degree the kinds of assembly and control processes needed to organize on a short term basis adaptive strategies for solving novel problems. The more complex and varied the sequence of novel problems, the more adaptive the processing needs to be. The Raven Progressive Matrices Test is perhaps the archetypical example of such a test (Snow, 1980, pp. 35-36).

According to this interpretation the most essential features of g tests are that they present novel and complex tasks. The novelty forces the examinee to find new ways of solving the tasks, and the complexity ensures that this is not simple: the examinee must always be prepared to find new modes of attack, and with greater complexity follows that the number of steps and intermediary results to keep track of rapidly increases.

Snow's interpretation of g is admittedly abstract and vague but at the present state of knowledge it does seem quite impossible to carry interpretations any further. What is important, however, is that an interpretation is couched in such terms that it may be developed into more specific formulations, and from this point of view the Snow approach does seem profitable: it relates directly to flourishing research on information processing, computer simulation, and artificial intelligence.

Among the second-order factors in the HILI-model Gc and Gv appear to be the most important ones. In discussing these factors it must be stressed that the factors labeled Gc and Gv in the HILI-model are not directly comparable with the factors with the same labels in the Cattell-Horn model. This is because in the HILI-model there is no variance from the g factor at the second-order level, while in the Cattell-Horn model the g variance is included in Gc and Gf. In order to separate these two ways of representing general intelligence, the residual factors, after g has been added, will be referred to as Gc' and Gv'.

It does appear most likely that Gc' is more or less identical with the v:ed factor of the Vernon model, even though there is as yet no empirical proof of this. Both factors, however, represent the verbal content area, with a strong leaning towards knowledge acquired through formal education, and both Vernon and Cattell-Horn mention tests of vocabulary as the best measure of the factor.

These factors also seem to be identical with a factor termed VEK (Verbal-Educational Knowledge) by Undheim (1981c). Undheim (1981b) argued that this factor may represent a rather narrow achievement factor:

> it may be related to opportunity, interest, and effort in verbal-educational achievement in school -- reflecting engaged time in school learning, in reading books more generally, reading newspapers and magazines, watching "educational" programs on TV, etc (Undheim, 1981b, p. 186)

Thus, Undheim sees Gc´ as being the accumulated result of choice of verbally oriented activities. Such a theoretical position comes close to the "transfer" theory proposed by Ferguson (1954), and is supported for example by findings that choice of educational and occupational tracks does affect the relative strength of verbal and spatial abilities (Balke-Aurell, 1981).

Snow (1980) also has proposed that Gc is the result of prior learning and argued that it:

> represents the long term accumulation of knowledge and skills, organized into functional cognitive systems by prior learning, that are in some sense crystallized as units for use in future learning. Since these are products of past education, and since education is in large part accumulative, transfer relations between past and future learning are assured. The transfer need not be primarily of specific knowledge but rather of organized academic learning skills. Thus Gc may represent prior assemblies of performance processes retrieved as a system and applied anew in instructional situations not unlike those experienced in the past... (Snow, 1980, p. 37)

A similar line of reasoning could be constructed to account for Gv.

In addition, however, to interpretations of Gc´ and Gv´ in terms of prior learning it may be that that these factors reflect differential processing requirements of verbal and figural information. Thus, in the research on brain laterality (e.g. Bock, 1973; Harris, 1975; Nebes, 1974) it has been established that there are two broad modes of processing, one associated with the left hemisphere and verbal information, and the other with the right hemisphere and figural information. The first mode of processing is described as analytic, linear, binary, serial or succesive, while the other mode is described as global, parallel, holistic, synchronous, simultaneous or continous. It could thus be that Gc´ and Gv´ express the facility with which these types of processing, respectively, are performed.

It thus seems that two different explanations of individual differences in Gc´ and Gv´ may be proposed: one that takes its starting point in differences in long term memory as a consequence of prior learning, and one that concentrates on the different processing characteristics of verbal and

figural information. These interpretations are of course not mutually exclusive and they may both be true. There may also be quite intricate relationships between the two mechanisms. Thus, small initial differences in proficiency in a certain type of processing may affect interests and preferences, such that large differences in acquired knowledge result. It is also conceivable that availability of a large knowledge base enhances and expediates the type of processes which operate on that knowledge base.

In the Cattell-Horn and HILI models the factors Gs and Gr are placed at the same level as Gc and Gv. In Vernon's model, in contrast, factors corresponding with these are placed at lower levels in the hierarchy. Thus, the kind of fluency factors being involved in Gr are by Vernon regarded as minor group factor below Gc. In comparison with Gv and Gc, however, these factors appear to be of lesser importance, and further empirical research may settle this minor difference between the hierarchical models.

There is little reason to discuss the wealth of primary factors at the lowest level in the HILI model. All the models considered here do include a smaller or larger set of narrow primary factors, and there is considerable overlap among the lists of factors, even though different labels may be employed (e.g. Guilford, 1972). In the HILI-model the primary factors represent the variance which is left after the variance from the higher-order factors has been partialed out. For many factors this is only a small fraction of the total variance, which residual in many cases may be of limited psychological interest.

3.3.2 Relationships with other models

The results which have been presented so far show that the HILI-model is compatible with each of the other models considered here. At the same time, however, the HILI-model goes beyond the previous models by taking into account individual difference variance at several levels of generality. The previous models may, therefore, be viewed as special cases of the HILI-model.

It is also interesting to compare the HILI-model with another representation of the organization of abilities, presented by Snow (1980, see also Snow, Lohman, Marshalek, Yalow & Webb, 1977). In this approach the techniques of multidimensional scaling and hierarchical cluster analysis are used instead of factor analysis. Application of these methods to correlation matrices for large numbers of tests typically yield a scatter, where each tests is represented as a point in two-dimensional space. At the very center of the scatter appear tests of Gf, while tests of Gv and Gc appear as clusters not

far from the center. In the more peripheral regions of the two-dimensional chart there appear clusters of tests which may be identified with primary factors.

Marshalek (1977) interpreted the degree of centrality of a test as reflecting the complexity of the processing involved, or the involvement of g. Marshalek also pointed out that this model of intelligence is compatible with hierarchical models, such as Vernon's model: The degree of centrality represents the level in the hierarchy, and the clustering of tests represents factors at different levels. It would thus seem that this model based upon multidimensional scaling is compatible with the HILI-model.

Sternberg (1980) recently presented a model of intelligence, in relation to which claims of generality have been made too. It is interesting, therefore, to compare the Sternberg model with the HILI model.

The "componential theory of intelligence" proposed by Sternberg is based on the concept of component, which is defined as "... an elementary information process that operates upon internal representations of objects or symbols" (Sternberg, 1980, p. 6). On the basis of function, components are classified into five different categories: meta-components, performance components, acquisition components, retention components, and transfer components. Meta-components "are higher order control processes that are used for executive planning and decision making in problem-solving" (p. 7), while performance components represent processes actually used in task performance.

Sternberg also classifies components according to level of generality into three categories: general components, class components and specific components. General components are processes used in all tasks within a given universe; class components are processes used within a sub-set of tasks; and specific components are used in the accomplishment of single tasks.

This classification of components is utilized in an assumed hierarchical organization of tasks. For each task in a hierarchy the same general components are used, and for each task different specific components are used. The level in the hierarchy at which a task is placed is determined by the class components: tasks at the lowest level each require one set of class components, while tasks at higher levels require all the class components of tasks at lower levels within the same branch of the hierarchy.

Sternberg confronted several models of the organization of human abilities with this componential conception of task performance. With respect to Spearman's Two Factor theory it was argued that the g factor comprises a set of general components that is common to a wide variety of tasks, while the s-factors correspond to specific components. It was, furthermore, argued

that the meta-components have a much higher proportion of general components among them, since for almost every task executive routines for planning and monitoring performance must be invoked. It was, thus, concluded that "individual differences in meta-componential functioning will be primarily responsible for the appearance of individual differences of a general nature" (p.10).

The Thurstone PMA's were by Sternberg interpreted to reflect individual differences in class components, while the correlation among the primary factors is accounted for by general components. As an example, Sternberg mentioned the I-factor, which was argued to involve a relatively small set of class components (i.e. inference, mapping, application, and justification).

The concepts of fluid and crystallized intelligence were also discussed. Tests of Gc were interpreted to reflect "the products of acquisition, retention and transfer components, whereas fluid ability tests seem to involve the execution of performance components".

Sternberg concluded that "factor theories of intelligence are all right almost. What this means is that almost all factor theories of intelligence are right in the sense of being special cases of a more general psychometric theory, but that they are not quite all right when considered in isolation. They need to be complemented by componential theories ..." (p. 12).

While there is no need to challenge the conclusion that componential theories are complementary to factorial models, it would not seem that the theory outlined by Sternberg is able to function as a super-theory, within which the different models of the structure of abilities are contained as special cases. This may be seen if the specific interpretations proposed by Sternberg are scrutinized.

Sternberg argues that the g factor in Spearman's Two Factor theory represents individual differences in meta-components; that the Thurstonian I-factor represents individual differences in a set of performance components; and that Gf reflects individual differences in the execution of performance components generally. It has been shown, however, that g is identical with Gf, and the empirical evidence also indicates that I is virtually identical with these higher-order factors. Sternberg thus proposes three different explanations for the same individual difference variance. Even though these explanation are not mutually exclusive, this indicates that the componential theory is much too loose to function as a general psychometric theory.

Even more important is the fact that while the factorial models identify and structure systematic sources of individual differences at different levels of

generality, the componential theory models performance on intellectual tasks. This very fundamental difference between the factorial and componential approaches is seen if the content of the hierarchies of the two models is scrutinized. In the componential theory the hierarchy is a hierarchy of tasks, while in the factorial approach it is a hierarchy of sources of individual difference variance. This difference in focus of attention makes the factorial and componential approaches complementary, but it also implies that the componential approach cannot provide a theory under which the factor-analytic models may be subsumed.

Before leaving comparisons with other models, it is interesting to consider the reasons why it has for so long been possible for several models to coexist, which, at the surface at least, provide so different accounts of the structure of human abilities.

Sternberg (1981a) attributed this to the fact that exploratory factor analysis is succesful in supporting almost any model. At a more concrete level, however, this may be explained by differences among the exploratory techniques, and particularly in the way they deal with variance from general factors. It is well known that while some techniques readily produce a strong general factor, others cannot even be forced to indicate the presence of a general factor.

It may easily be demonstrated (e.g. Gustafsson et al., 1981; Humphreys, 1979) that when Multiple Factor analysis is used with orthogonal rotation, the general factor is "rotated away", by being represented as small positive loadings in all factors. However, in interpretations of factor analytic findings, loadings lower than .30 are rarely attended to, and often not even presented. It may thus be claimed that orthogonal rotations to simple structure are quite deceptive in the presence of a general factor.

If an oblique rotation is carried out, the general factor is represented as the correlation among the factors. There are two problems inherent in oblique rotations, however. One is that there are almost always small positive loadings scattered in the matrix, which cause the true correlation among factors to be underestimated. The other problem is that most oblique rotational methods allow the researcher to determine the degree of obliqueness of the solution: In the Promax method (Hendrickson & White, 1964) this is governed by the parameter k; in the indirect oblimin method (e.g. Harman, 1967, pp. 325-326) obliqueness is governed by the parameter gamma, and so on. Oblique rotational methods can, therefore, not provide "objective" empirical information on the amount of actual correlation between factors, and thereby not information on the importance of higher-order factors. In confirmatory factor analysis, however, all these problems are avoided.

3.3.3 The nature of the HILI-model

The HILI-model has been presented as a general model of the structure of human abilities, of which most other models may be viewed as special cases. This should not be interpreted to mean, however, that the model has reached its final form. On the contrary, it should be viewed as an open-ended model to be elaborated upon.

It may, in fact, be more correct to talk about a class of HILI-models, than about the HILI-model. Models belonging to this class share the features that they are hierarchical, are based upon LISREL, and include a set of "basic" factors, such as g, Gc and Gv. The number of levels in the hierarchy may differ from model to model, however, as may the domains of tasks which are sampled. Thus, when a certain area is represented by few tests there will necessarily be few levels in that branch of the hierarchy, as was the case for Gc in the reanalysis of the Undheim (1978) study. When an area is investigated in great detail, as Fredriksen (e.g. 1980) has done with reading, for example, it may be possible to add levels both below and above what is taken to be the primary level in the HILI-model as presented in Figure 4.

The fact that the HILI-model is formulated within the framework of LISREL not only brings the advantages of confirmatory maximum likelihood factor analysis. LISREL is a general data-analytic system which allows formulation of structural equation models in which latent variables are related to one another. This implies that it is possible to formulate models in which the factors at the different levels of the HILI-model are used as independent and/or dependent variables. Such models have several advantages in comparison with for example ordinary multiple regression analysis: they are parsimonious; problems associated with errors of measurement are avoided; and there is hope that factors may remain invariant from one model to another.

4 IMPLICATIONS OF THE MODEL

As is evident from the fact that the HILI-model is a synthesis of previously existing models, not much in it is new. In spite of this, however, it does carry implications both for how the structure of abilities is looked at, and for testing practices. In this section some of these implications are outlined.

4.1 Implications for testing

The technology of mental measurement has, for good and for bad, had major practical impact, and ever since the Binet and Simon instrument was introduced tests have been used for purposes of diagnosis, selection and classification.

It may be noted, though, that the technology of testing has remained relatively unaffected by the developments of factor-analytically based models of ability. Thus, in spite of the fact that none of the "modern" models includes a general factor, tests of general mental ability are frequently used in assessments. These tests are heterogeneous tests in the Binet-Simon-Terman tradition, and they are quite similar to the early tests:

> Ability tests have remain about the same since 1920 ... The practical tests of today differ from the tests of 1920 as todays's automobiles differ from those of the same period: more efficient and more elegant, but operating on the same principles as before (Cronbach, 1960, p. 159).

This kind of tests assumes the existence of a factor of general ability. Considering, however, that there has neither been a theory to account for performance on the tests, nor a place for general ability in the influential models of ability. it is quite amazing that this type of tests has been used at all. The reason for this must sought in the fact that they have been demonstrated to have practical utility.

The HILI-model does revitalize the Spearman g factor and places it at the most prominent place in the hierarchy. From the "principle of indifference of the indicator" it follows that a sufficiently long and heterogenous test is virtually perfectly correlated with the g factor. The HILI-model thus provides a rationale for the established practice of assessing general ability with heterogenous tests, even though the actual amount of correlation between such tests and g has yet to be determined. It should also be stressed, of course, that in many cases a test of general mental ability may not provide the information that is sought (see below).

In another type of frequently used tests the items are organized in subscales. Examples of such tests are the Wechsler scales and the British Intelligence Test (Warburton, 1970). However, even though the subscores are available they are often summed to yield an overall score, or partial sums. When an overall score is obtained it provides an assessment of general ability, in the same way as any heterogenous test does.

The meaning of partial sums of sub-test scores is of course dependent upon which particular scales enter the sum. Often, however, a contrast is made between Verbal and Performance areas. These areas of competence appear to come quite close to the higher-order factors Gc and Gv, respectively, even though the actual amount of similarity must be established empirically for each test. Assuming, however, that the amount of correspondence is substantial it would seem that this practice is rationalized by the HILI-model; the Verbal and Performance scores may be viewed as crude estimates of the second-order factors.

It thus seems that the HILI-model is compatible with much established testing practice. The non-hierarchical models based upon Multiple Factor analysis would, in contrast, seem to carry the implication that such higher-order constructs as general ability or Performance IQ should not be assessed at all.

The HILI-model goes, however, much beyond the established practice. Most important is the fact that the hierarchical structure of the model makes it a general and versatile tool for describing individual differences. Thus, with a hierarchical model it is possible to select not only the appropriate areas of competence for assessment, but also the appropriate level of detail in the assessments:

> Microanalysis is neither more or less correct than gross analysis; the size of the bundle into which abilities are tied should be adjusted to the theoretical and practical context. At times elaboration is needed to communicate an elaborate thought. At other times, it confuses. (Cronbach & Snow, 1977, p. 154).

Suppose, to take a few simplistical examples, that the purpose is to predict school achievement without regard to line of study. In such a case it would most likely suffice to measure g and Gc. If, however, we for guidance purposes want to predict achievement in verbal and technical lines of study, we would also want to differentiate between Gv and Gc. Should we want to make even finer differential predictions, such as between different contents in a technical education, we could make measurement distinctions further down in the hierarchy, i.e. at the primary level, or at the level of specific tests.

It is obvious that to estimate all factors at all levels with sufficient accuracy, huge test-batteries would be required. This amount of information is rarely needed, however, and to estimate just one or a few of the higher-order factors a rather limited battery should suffice. Thus the g factor should be well represented in a test-battery consisting of one test of the I-factor, one V-test and one Vz-test. It is a rather trivial task to

determine appropriate batteries, and coefficients for estimating individual factor scores, for different applications.

4.2 Implications for teaching

The HILI-model in itself does not carry any implications for how to take into account individual differences in teaching. However, the question how to adapt instruction to accomodate individual differences in ability is being studied in research on aptitude-treatment interactions (ATI, cf Cronbach & Snow, 1977). The basic premise of this line of research is that no educational treatment is optimal for everyone, but that different subgroups of pupils may profit from different teaching methods. It may also be shown (cf Cronbach & Gleser, 1956) that any attempt at differential treatment of pupils, such as streaming and different kinds of individualization, is based on the assumption of an interaction between aptitude and treatment.

The most basic design of an ATI-study involves one or more aptitude variables, two educational treatments to which the pupils are randomly assigned, and one or more outcome variables. It is then determined, most commonly with multiple regression (MR) analysis, in which treatment subjects with different levels of performance on the aptitude variables achieved the best results.

So far ATI-research has not been very succesful in producing a knowledge base for how to adapt instruction to individual differences (Cronbach, 1975; Cronbach & Snow, 1977), even though some tentative generalizations have been reached (Snow, 1977). While this is not the place to discuss all the problems intrinsic to ATI-research (cf Cronbach, 1975; Cronbach & Snow, 1977; Gustafsson, 1981; Snow, 1977, 1980). it is obvious that the confusion concerning the structure of abilities is to a certain extent responsible for the lack of success. Thus, researchers have selected and interpreted aptitude variables within different frames of reference, which in turn has caused great problems assembling and integrating the findings.

In much ATI-research primary factors of ability have been the starting point. Aptitudes and treatment have, however, been matched on a rather superficial basis, as when it is hypothesized that verbal types of instruction would benefit pupils with high verbal ability, and that spatial-pictorial types of instruction would benefit pupils with high spatial ability (cf. Gustafsson, 1976). This research has not been productive of any strong and generalizable findings, which may be due to the fact that lower-order rather than higher-order factors have been concentrated upon.

Another reason why ATI-research has tended to produce conflicting and inconclusive findings may be that the MR-technique is based upon assumptions which are rarely fulfilled. Thus, in MR it is assumed that there are no errors of measurement in the aptitude variables, and that variables simultaneously entered into an analysis do not measure a common underlying factor. However, almost all ability tests are more or less unreliable, and it very frequently happens that variables with a high true correlation are analyzed together. It can easily be demonstrated that such violations of the assumptions may cause MR to produce misleading results (cf Gustafsson & Lindström, 1979).

The LISREL-technique may, however, provide a solution to these technical problems. Through basing the analysis on latent, rather than observed variables, LISREL avoids all the problems associated with MR. What is even more important, however, is the fact that the HILI-model may be used as a hierarchical measurement model within LISREL. In this way relations may be determined between factors at different levels in the model and learning outcomes, which relations may be tested for equality in different treatments.

Such an approach to the design and analysis of ATI-studies brings several advantages:

- The model is encompassing enough to cover a wide range of aptitude variables, which makes it rather generally useful. This reduces the problem of "translating" aptitude variables from one frame of reference to another when results from different studies are compared and integrated.

- The fact that the model is hierarchical makes it possible to formulate extremely parsimonious models for the relations with other variables, by invoking first the g factor, and then invoking only as many of the lower-order factors as may be necessary (cf Gustafsson et al., 1981).

- Even in those cases when so few variables are measured that the model cannot be used in full, the hierarchical approach and mode of thinking may be utilized, and often it should be possible to interpret the factors within the framework of the HILI-model. For example, if in a study interest is centered on the g factor a selection of three or four tests regresenting Gc, Gf and Gv may yield one common factor. This factor should come very close to the third-order g in the HILI model. The results in such a study may be compared to the results in another study with a much larger test battery, even though it will of course not be possible to separate error variance, test-specificity, and the residuals of primary and second-order factors.

Since the LISREL-technique is rather new and quite complex, its use in

ATI-research is best presented with an example.

4.2.1 An example of the HILI-model in ATI-research

The study to be presented briefly here is presented in full by Gustafsson (1982). The purpose of the study was to investigate ATI-effects between aptitude variables and the treatment dimensions reading/listening and pictures/no pictures.

In laboratory studies it has been shown that when subjects listen to messages they acquire spatial content more efficiently than when they read the messages, which has been interpreted as an effect of suppression of visualization by reading (e.g. Brooks, 1967). This finding raises the questions whether modality of presentation may be shown to have similar effects also for meaningful material, and whether such suppression effects are moderated by abilities, such as g and Gv.

Research on the effects of pictures in instructional materials has yielded a rather mixed pattern of results, both when it comes to main effects and interactions with abilities. However, should the hypothesis of suppression of visualization by reading be true, it would seem likely that reading hampers interpretation and utilization of pictorial information as well.

In the study the two treatment dimensions were crossed to yield four treatments: reading an unillustrated text (READVERB), reading an illustrated text (READPICT), listening to a presentation via taperecorder of the text material (LISTVERB), and listening to a presentation via taperecorder with illustrations presented as slides (LISTPICT). Each treatment group consisted of some 100 subjects from the 5th grade.

The material taught dealt with the heart and the flow of blood in the body. This subject matter was chosen because it was felt that it would in part be of a spatial character and that visualization processes thus would be helpful in acquiring it. Immediately after the instruction the subjects were given two post-tests: one verbal and one pictorial. From the items in these post-tests several different scales were constructed. One of these covered the learning of verbal types of information and is of no particular interest in this context. Two others dealt with spatial types of information (i.e. to and from which parts of the heart blood flows in relation to different parts of the body), one with verbal items (V-SPAT) and the other with pictorial items (P-SPAT). These two scales will be concentrated upon in the analysis.

Aptitude variables in the study were four tests: the vocabulary test Opposites (Op); the visualization test Metal Folding (MF); the reasoning test Number Series (NS); and a paired associates (PA) learning test, involving words and pictures. These tests may be taken to represent the primary factors V, Vz, I and Ma, respectively.

With such a limited battery it is obviously impossible to separate factors of different orders. The best that may be achieved with only four observed variables is a model with one latent variable, which with the present heterogeneous collection of tests is most likely to represent the g factor. However, by entering the observed variables as half-tests, for example by splitting the tests into odd and even items, one more level may be added to the model, which level represents the true variance in the observed variables. For simplicity we will designate these latent variables with the labels of the second-order factors in the HILI-model, i.e. Op will be referred to as Gc, MF as Gv, and NS will be labelled Gf. Such an equivalencing of test variance with second-order factors is of course from a strict point of view incorrect. However, these tests load very strongly in their respective primary factors, which in turn load very strongly in their respective second-order factors (cf Gustafsson et al., 1981), so it may be taken as a fairly good approximation.

The measurement model is shown in Figure 5.

Insert Figure 5 about here

It has been constructed as a hierarchical model in which the residuals of the lower-order factors are taken to be latent variables, so that they are available for purposes of prediction. Estimating this model from the covariance matrices for each of the four treatment groups a good fit was obtained (chi-square=133.0, df=128). This model may thus be taken as a good representation of the relations among the tests in the battery, and it may be concluded that the treatment groups are samples from the same population.

4.2.2 Results

The model used in the first step of the analysis is shown in Figure 6.

Insert Figure 6 about here

This model is quite simple, the observed outcome variables being regressed onto the g factor of the hierarchical measurement model. It may of course be that this model does not fit well within the treatment groups. Should this be the case it is most likely due to the fact that the lower-order factor contribute to the prediction of one or both of the outcome variables, and then these relations should be invoked as well. Before presenting the results from the ATI-analysis, however, the results from analyses of main effects will be presented.

According to the hypothesis of suppression of visualization by reading we would expect a higher level of performance in the treatments which do not involve reading, i.e. the LISTVERB and LISTPICT treatments. Performance was indeed slightly higher in the LISTVERB than in the READVERB treatment, but this effect failed to reach statistical significance. It was, however, found that the level of performance was significantly higher in the LISTPICT treatment than in the other three treatments. This finding provides some support for the hypothesis of suppression of picture interpretation by reading.

In the analysis of relations between aptitudes and outcomes within treatments it was found that in every treatment the regression on g was significant, even though it appeared to vary in strength. It also was found that in some of the treatments there were significant relationships between Gv′ and the outcomes. Thus, for the analysis of ATI-effects the model included g and Gv′ as predictors. The estimates of the within-treatment regression coefficients are presented in Table 1.
As may be seen from the Table the regression of the outcomes on g was steeper in the treatments involving listening than in the treatments involving reading. The contrast between listening/reading treatments was statistically significant (chi-square=5.49 and 9.13 with 1 df for V-SPAT and P-SPAT, respectively), but the small observed differences between the LISTVERB and LISTPICT treatments were not significant.

As may also be seen from the Table the regression of V-SPAT on Gv′ was

---

Insert Table 1 about here

---

steeper in the READVERB and READPICT treatments than in the other two. Again the contrast between treatments involving reading/not involving reading was significant (chi-square=4.17, df=1).

The final finding worth mentioning is that the P-SPAT outcome had a highly significant regression on Gv´ in the READPICT treatment, and no relationship with Gv´ in any of the other treatments.

4.2.3 Discussion of the results

It should first of all be pointed out that the data analyzed here have also been analyzed with MR. This analysis did not, however, yield any interpretable findings, which is most likely due to the fact that the present data severely violate the assumptions of MR.

It should also be mentioned that the LISREL-analysis has also been carried one step further, by division of the treatment groups according to sex. In that analysis several higher-order interactions with sex were disclosed, but it would carry to far to discuss them here (see instead Gustafsson, 1982).

Almost all the ATI-effects found involve the treatment factor reading versus listening and the aptitude variables g and Gv´. It was expected that subjects with a high level on Gv´ would perform especially well in the treatments involving listening, because in these treatments they would have the opportunity to use visualization processes. Instead g was found to be particularly conducive to achievement in these treatments.

This may be because in these treatments the proper sequencing and interrelating of processes is important: The incoming verbal information must be decoded, and the spatial type of content must be further dealt with in visualization processes, and while this is done new information arrives. This may thus put high demands on the efficiency of the control and assembly processes hypothesized to be at the core of g.

Quite surprisingly Gv´ was found to be more highly related to outcome in the

reading than in the listening treatments. This finding may, perhaps, be due to the high-Gv´ pupils being able to carry out visualization even under the influence of suppression effects from reading.

It was also found that the P-SPAT outcome regressed more steeply on Gv´ in the READPICT treatment than in the other treatments. This seems to be one of the few findings in support of the hypothesis that high-spatial pupils in particular profit from pictorials. The effect was, however, restricted to the pictorial type of outcome measure, which may be due to the fact that Gv´ is of importance in an illustrated treatment only when acquisition of the illustrations themselves is beneficial for achievement (cf. Gustafsson, 1976).

5 CONCLUDING REMARKS

In the present paper a model of the organization has been presented, which model in a certain sense is more general than previous models. Given, however, that the question of structure has been given an answer, questions about the meaning and interpretation of the dimensions of ability come into focus. The HILI-model is, of course, just as unable to answer such questions as any other structural model.

However, to get information upon which to base interpretations of the nature of different dimensions of intelligence one may study relations between these and other variables. The ATI-study presented above purports to show that the LISREL-framework is extremely well suited to conduct such studies. The major importance of the HILI-model may, therefore, reside in the possibilities to study with precision relations between ability factors of different degrees of generality, and other variables related to learning and cognition.

Table 1. Estimated within-treatment regression coefficients

Regression of	on	Treatment READVERB	READPICT	LISTVERB	LISTPICT
V-SPAT	G	.16	.15	.43	.54
P-SPAT	G	.20	.25	.55	.70
V-SPAT	Gv′	.12	.13	.03	.01
P-SPAT	Gv′	.02	.18	.00	.03

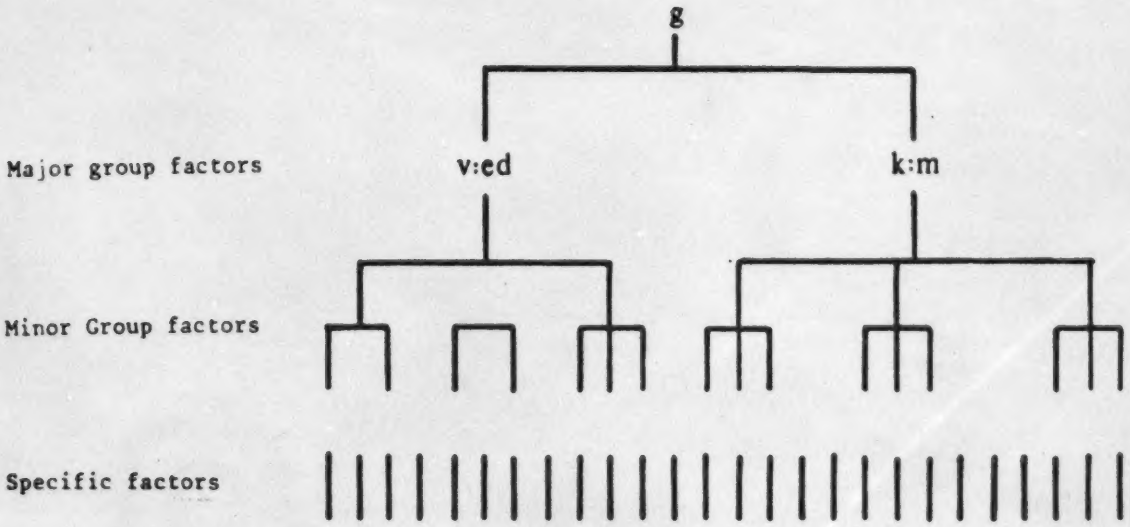

Figure 1. The Vernon model.

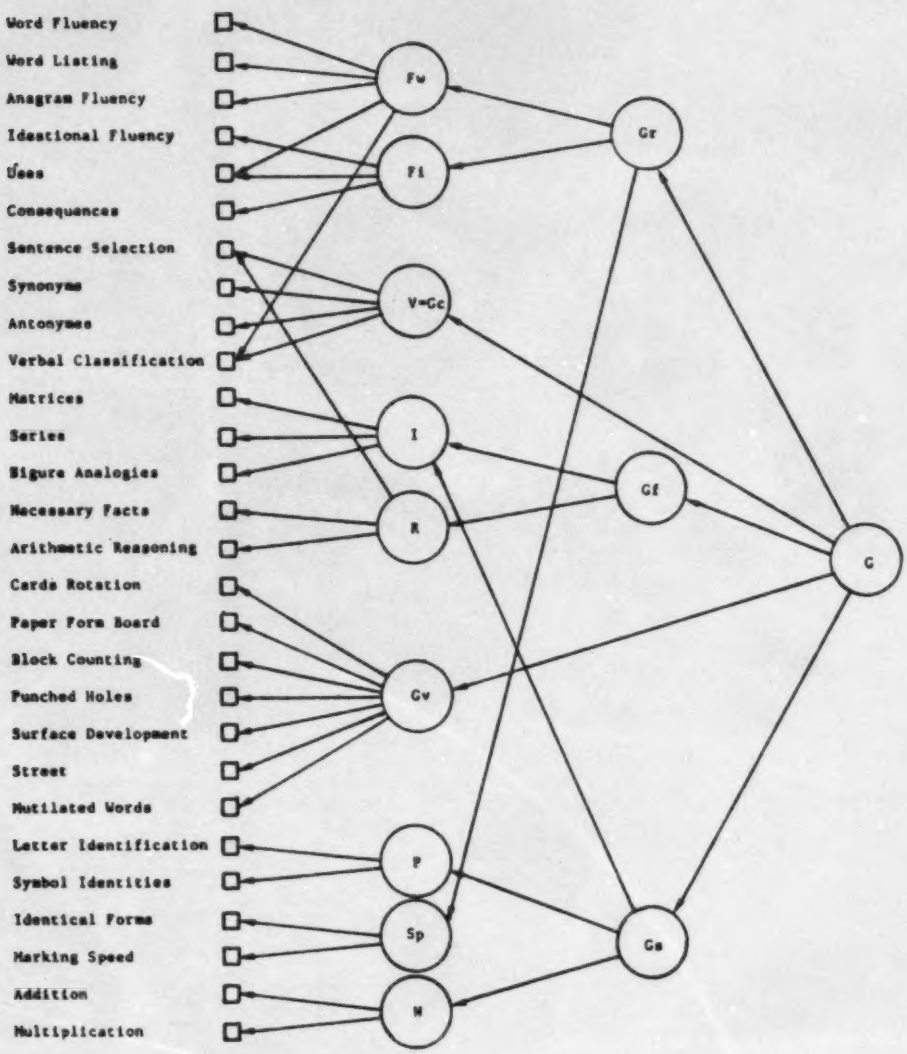

Figure 2. The final model in the reanalysis of the Undheim (1978) study.

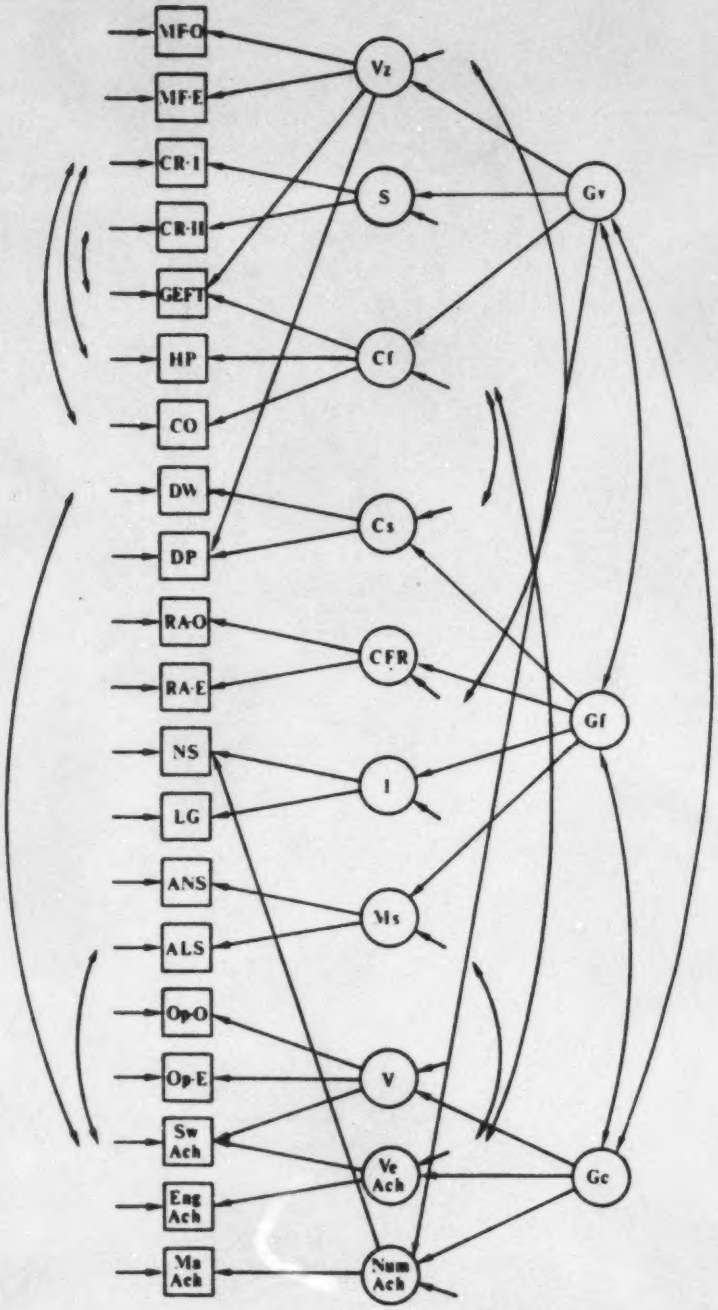

Figure 3. The second-order model in the Gustafsson et al. (1981) study.

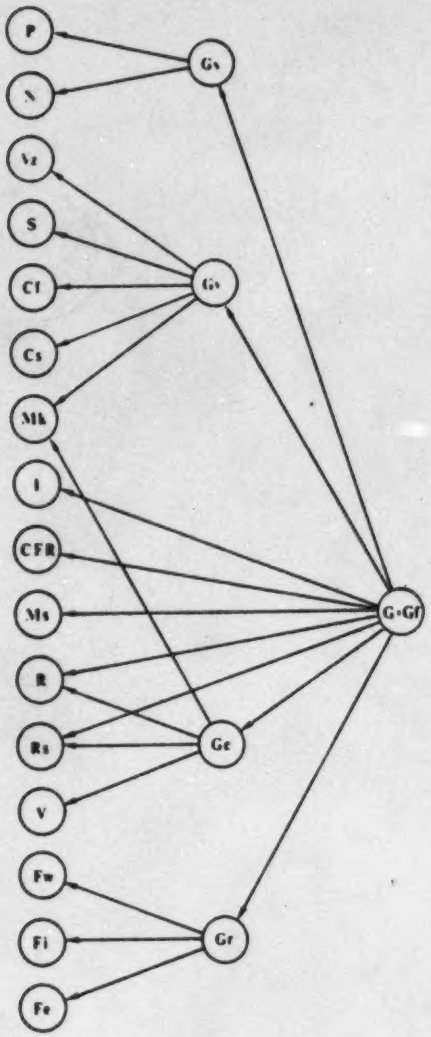

Figure 4. An hypothesized hierarchical model for the most well established primary factors.

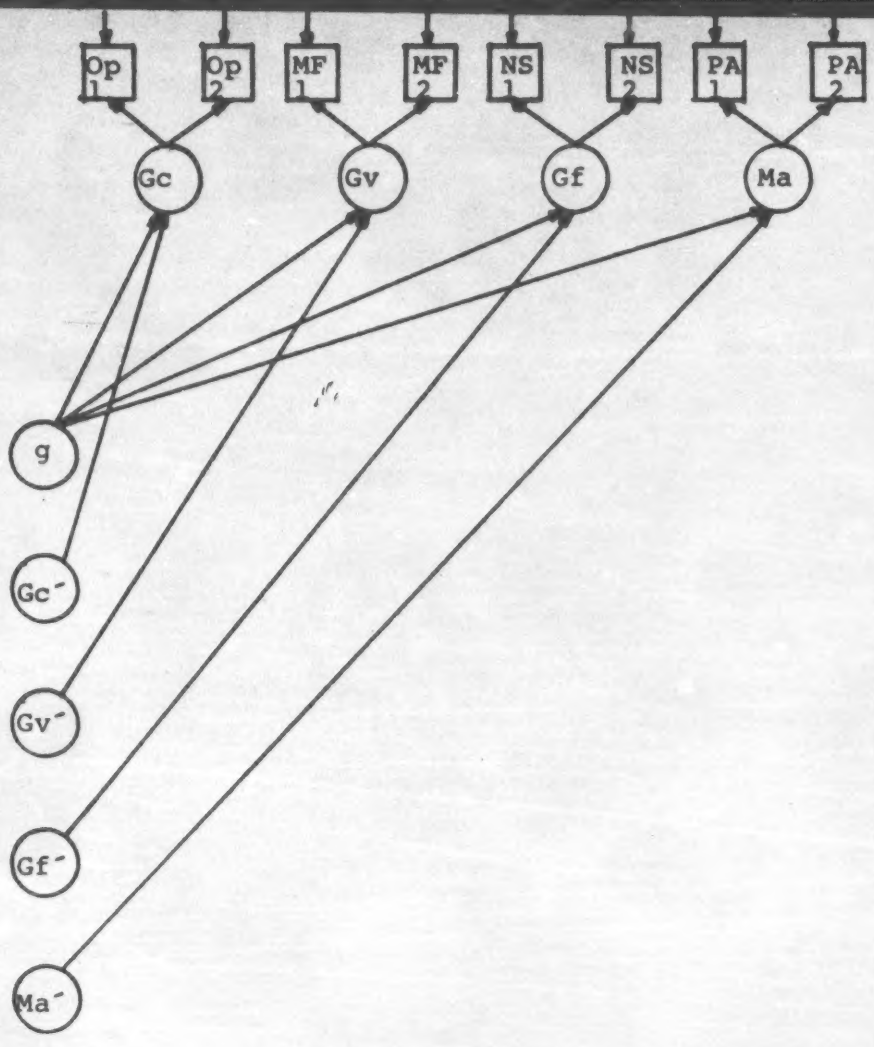

Figure 5. The measurement model for the aptitude variables in the Gustafsson (1982) study.

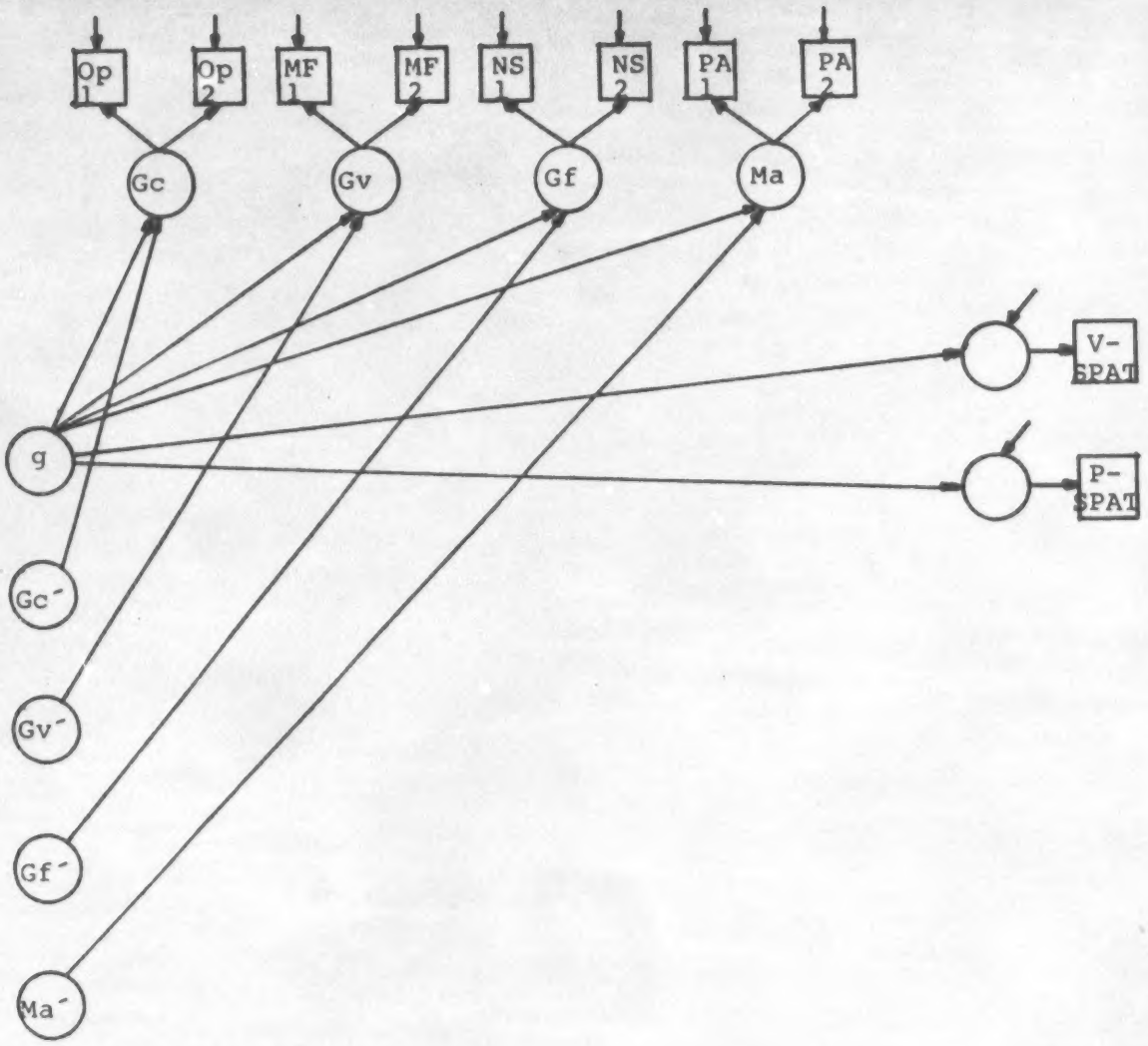

Figure 6. The starting model for the ATI-analyses in the Gustafsson (1982) study.

REFERENCES

Balke-Aurell, G. Changes in ability factors as related to educational and occupational experience. Göteborg: Acta Universitatis Gothoburgensis, 1981.

Bock, R.D. Word and image: Sources of the verbal and spatial factors in mental test scores. Psychometrika, 1973, 38, 437-457.

Brooks, L.R. The suppression of visualization in reading. Journal of Experimental Psychology, 167, 19, 289-299.

Burt, C.L. The factors of the mind: An introduction to factor analysis in psychology. New York: MacMillan, 1941.

Burt, C.L. The structure of the mind: A review of the results of factor analysis. British Journal of Educational Psychology, 1949, 19, 100-111; 176-199.

Cattell, R. B. A culture-free intelligence test, I. Journal of Educational Psychology, 1940, 31, 161-179.

Cattell, R.B. Theory of fluid and crystallized intelligence: A critical experiment. Journal of Educational Psychology, 1963, 54, 1-22.

Cattell, R.B. Abilities: Their structure, growth, and action. Bostopn: Houghton Mifflin, 1971.

Cronbach, L.J. Essentials of psychological testing (2nd ed). New York: Harper and Row, 1960.

Cronbach, L.J. Beyond the two disciplines of scientific psychology. American Psychologist, 1975, 30, 116-127.

Cronbach, L.J., & Gleser, G.C. Psychological tests and personnel decisions. Urbana: University of Illinois Press, 1957.

Curtis, M.E. & Glaser, R. Changing conceptions of intelligence. Review of Research in Education, 1981, 9, 111-148.

Ferguson, G.A. On learning and human ability. Canadian Journal of Psychology, 1954, 8, 95-112.

French, J.W. The description of aptitude and achievement tests in terms of rotated factors. Psychometric Monographs, 1951, No. 5.

French, J.W., Ekstrom, R.B., & Price, L.A. Manual for kit of reference tests for cognitive factors. Educational Testing Service, Princeton, New Jersey, 1963.

Fredriksen, J.R. Component skills in reading: Measurement of individual differences through chronometric analysis. In R.E. Snow, P.-A. Federico and W. W. Montague (Eds) Aptitude, learning, and instruction, Volume I: cogntive process analyses of aptitude. Hillsdale, New Jersey: Lawrence Erlbaum, 1980. po951 d

Galton, F. Inquiries into human faculty and its development. London: MacMillan, 1883.

Guilford, J.P. The nature of human intelligence. New York: Mc Graw-Hill, 1967.

Guilford, J.P. Thurstone's primary mental abilities and structure-of-intellect abilities. Psychological Bulletin, 1972, 77, 129-143.

Guilford, J.P., & Hoepfner, R. The analysis of intelligence. New York: McGraw-Hill, 1971.

Gustafsson, J.-E. Verbal and figural aptitudes in relation to instructional methods. Studies in aptitude-treatment interactions. Göteborg: Acta Universitatis Gothoburgensis, 1976.

Gustafsson, J.-E. Matching aptitudes and treatments: The ATI-paradigm in research on individuualization of instruction. In N Sövik, H.M. Eikeland & A. Lysne (Eds) On individualized instruction. Oslo: Universitetsförlaget, 1981.

Gustafsson, J.-E. Visualization processes in learning as a function of method of presentation and individual differences. Manuscript in preparation, 1982.

Gustafsson, J.-E. & Lindström, B. Analyzing ATI data by structural analysis of covariance matrices. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1979.

Gustafsson, J.-E., Lindström, B. & Björck-Åkesson, E. A general model for the organization of cognitive abilities. Reports from the Department of Education, University of Göteborg, 1981:06.

Harman, H. Modern factor analysis. Chicago: University of Chicago Press, 1967.

Härnqvist, K. Manual till DBA. (Manual for the DBA battery). Stockholm: Skandinaviska Testförlaget, 1968.

Harris, L.J. Neuropsychological factors in the development of spatial skills. In J. Eliott & N.J. Salkind (Eds) Children's spatial development. Springfield, Illinois: Charles C. Thomas, 1975.

Horn, J. L. Fluid and crystallized intelligence: a factor analytic study of the structure among primary abilities. Unpublished doctoral dissertation, University of Illinois, 1965.

Horn, J. L. Organization of abilities and the development of intelligence. Psychological Review, 1968, 79, 242-259.

Horn, J. L. State, trait and change dimensions of intelligence. British Journal of Educational Psychology, 1972, 42, 159-185.

Horn, J. L. Personality and ability theory. In R.B. Cattell & R.M. Dreger (Eds) Handbook of personality theory. Washington: Hemisphere, 1977.

Horn, J.L. Concepts of intellect in relation to learning and adult development. Intelligence, in press.

Horn, J. L. & Cattell, R.B. Refinement and test of the theory of fluid and crystallized intelligence. Journal of Educational Psychology, 1966, 57, 253-270.

Humphreys, L.G. The organization of human abilities. American Psychologist, 1967, 17, 475-483.

Humphreys, L.G. Critique of "Theory of fluid and crystallized intelligence: A critical experiment." Journal of Educational Psychology, 1967, 58, 129-136.

Humphreys, L.G. The construct of general intelligence. Intelligence, 1979, 3, 105-120.

Jöreskog, K.G. A general approach to confirmatory maximum likelihood factor analysis. Psychometrika, 1969, 34, 183-202.

Jöreskog, K.G. A general method for analysis of covariance structures. Biometrika, 1970, 37, 239-251.

Jöreskog, K.G. A general method for estimating a linear structural equation system. In A.S. Goldberger & O.D. Duncan (Eds) Structural equation models in the social sciences. New York: Seminar Press, 1973.

Jöreskog, K.G., & Sörbom, D. LISREL IV. A general computer program for estimation of linear structural eqution systems by maximum likelihood methods. University of Uppsala, Department of Statistics, 1978.

Jöreskog, K.G., & Sörbom, D. LISREL V. Analysis of linear structural relationships by maximum likelihood and least squares methods. University of Uppsala, Department of Statistics, 1981.

Kelley, T.L. Crossroads in the mind of man. Stanford, California: Stanford University Press, 1928.

Lohman, D. Spatial ability - A review and reanalysis of the correlational literature. Technical Report no. 8, Aptitude Research Project, School of Education, Stanford University, 1979.

Marshalek, B. The complexity dimension in the radex and hierarchical models of intelligence. Paper presented at the American Psychological Association Convention, San Francisco, 1977.

McNemar, Q. Lost: our intelligence? Why? American Psychologist, 1964, 19, 871-882.

Nebes, R.D. Hemispheric specialization in commisurotomized man. Psychological Bulletin, 1974, 81, 1-14.

Pawlik, K. Concepts and calculations in human cognitive abilities. In R.B. Cattell (Ed) Handbook of multivariate experimental psychology. Chicago: Rand McNally, 1966.

Resnick, L.B. (Ed) The nature of intelligence. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1976.

Smith, I.M. Spatial ability. Its educational and social significance. London: University of London Press, 1964.

Snow, R.E. Research on aptitudes: A progress report. In L.S. Shulman (Ed) Review of Research in Education, Vol. 4. Itasca, Illinois: Peacock, 1977.

Snow, R.E. Theory and method for research on aptitude processes. Intelligence, 1978, 2, 225-278.

Snow, R.E. Aptitude processes. In R.E. Snow, P.A. Federico & W. Montague (Eds) Aptitude, learning, and instruction: Cognitive process analysis. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1980.

Spearman, C. The proof and measurement of association between two things. American Journal of Psychology, 1904a, 15, 72-101.

Spearman, C. General intelligence objectively determined and measured. American Journal of Psychology, 1904b, 15, 201-293.

Spearman, C. The abilities of man. London: MacMillan, 1927.

Sternberg, R.J. Intelligence, information processing, and analogical reasoning. The componential analysis of human abilities. Hillsdale, N.J.: Erlbaum, 1977.

Sternberg, R.J. Factorial theories of intelligence are all right almost. Educational Researcher, 1980, 9(8), 6-13.

Sternberg, R.J. Intelligence and nonentrenchment. Journal of Educational Psychology, 1981a, 73, 1-16.

Sternberg, R.J. Nothing fails like success: The search for an intelligent paradigm for studying intelligence, Journal of Educational Psychology, 1981b, 73, 142-155.

Svensson, A. Relative achievement. School performance in relation to intelligence, sex and home environment. Stockholm: Almqvist & Wiksell, 1971.

Thurstone, L.L. Multiple factor analysis. Psychological Review, 1931, 38, 214-230.

Thurstone, L.L. Primary mental abilities. Psychometric Monographs, No. 1, 1938.

Thurstone, L.L. An experimental study of simple structure. Psychometrika, 1940, 5, 153-168.

Thurstone, L.L. A factorial study of perception. Chicago: University of Chicago Press, 1940.

Thurstone, L.L. Multiple factor analysis. Chicago: University of Chicago Press, 1947.

Thurstone, L.L., & Thurstone, T.G. Factorial studies of intelligence. Psychometric Monographs, No. 2, 1941.

Undheim, J.O. Ability structure in 10-11 year-old children and the theory of

fluid and crystallized intelligence. Journal of Educational Psychology, 1976, 68, 411-423.

Undheim, J.O. Broad ability factors in 12- to 13 year-old chldren, the theory of fluid and crystallized intelligence, and the differentiation hypothesis. Journal of Educational Psychology, 1978, 70, 433-443.

Undheim, J.O. On intelligence I: Broad ability factors in 15-year-old children and Cattell's theory of fluid and crystallized intelligence. Scandinavian Journal of Psychology, 1981a, 22, 171-179.

Undheim, J.O. On intelligence II: A neo-Spearman model to replace Cattell's theory of fluid and crystallized intelligence. Scandinavian Journal of Psychology, 1981b, 22, 181-187.

Undheim, J.O. On intelligence IV: Toward a restoration of general intelligence. Scandinavian Journal of Psychology, 1981c, 22, 251-265.

Vernon, P.E. The structure of human abilities. London: Methuen, 1950.

Vernon, P.E. Ability factors and environmental infuences. American Psychologist, 1965, 20, 723-733.

Vernon, P.E. Intelligence and cultural environment. London: Methuen, 1969.

Warburton, F. The British intelligence scale. In W.B. Dockrell (Ed.) On intelligence. Toronto: Ontario Institute for Studies in Education, 1970, pp. 71-98.

Wissler, C. The correlation of mental and physical traits. Psychological Monographs, 1901, 3, No. 16.

Witkin, H.A., Oltman, P.K., Raskin, E. & Karp, S.A. A manual for the Embedded Figures test. Palo Alto: Conculting Psychologists Press, 1971.

Reports from THE DEPARTMENT OF EDUCATION UNIVERSITY OF GÖTEBORG

1980:01 Johansson, Britt: Need of knowledge in nursing and demand for knowledge in nursing education. July 1980

1980:02 Dahlgren, Lars-Owe och Franke-Wikberg, Sigbrit: Social Structure of Society through the Eyes of University Students. October 1980

1981:01 Svensson, Lennart: The concept of study skill(s). Paper presented at the Sixth International Conference on Improving University Teaching. Lausanne, Switzerland, July 9-12, 1980. March 1981

1981:02 Gustafsson, Jan-Eric: An introduction to Rasch's measurement model. March 1981

1981:03 Alexandersson, Claes: Amadeo Giorgis empirical phenomenology. March 1981

1981:04 Åsberg, Rodney: How to understand cognitive differences from a cross-cultural perspective? August 1981

1981:05 Reuterberg, Sven-Eric och Svensson, Allan: Financial aid and class bias in higher education. October 1981

1981:06 Gustafsson, Jan-Eric, Lindström, Berner & Björck-Åkesson, Eva: A general model for the organization of cognitive abilities. December 1981

1982:01 Englund, Tomas & Svingby, Gunilla: Scientification of teaching - a concept to be questioned. February 1982.

1982:02 Dahlgren, Lars-Owe: Higher education - Impact on students. June 1982

1982:03 Gustafsson, Jan-Eric: New models of the structure of intellectual abilities: Implications for testing and teaching practice. October 1982.